



Αριθμός Πρόσκλησης- ΔΥΠΕΚΕ - 51823042

Αντικείμενο: «Μοναδικοποίηση εγγραφών πελατειακής βάσης σε επίπεδο Πελάτη ευρέως γνωστό ως Entity Resolution ή Golden Record Analysis σε αρχιτεκτονικό περιβάλλον MS Azure Cloud»

ΠΕΡΙΓΡΑΦΗ ΥΠΗΡΕΣΙΩΝ

## Συνοπτική Περιγραφή του έργου

Αντικείμενο : Μοναδικοποίηση εγγραφών πελατειακής βάσης σε επίπεδο Πελάτη ευρέως γνωστό ως Entity Resolution ή Golden Record Analysis σε αρχιτεκτονικό περιβάλλον MS Azure Cloud

Η προετοιμασία και ο καθαρισμός των master δεδομένων της πελατειακής βάσης της ΔΕΗ αποτελεί απαραίτητη προϋπόθεση για την ιστορική μετάπτωση των δεδομένων στην ενοποιημένη βάση της μελλοντικής πλατφόρμας CRM της επιχείρησης. Μετά την ολοκλήρωση του έργου του καθαρισμού και επιβεβαίωσης των δεδομένων πελατείας της ΔΕΗ, προέκυψε η ανάγκη για πρόσθετες τεχνικές εργασίες περαιτέρω εμπλουτισμού της master πληροφορίας πελατειακής βάσης.

Ειδικότερα, χρειάζεται να προστεθούν στοιχεία gender, στατιστικά στοιχεία ορθότητας της υφιστάμενης μεθοδολογίας ομαδοποίησης, προσθήκη κάποιων περαιτέρω πεδίων σε αυτήν, περαιτέρω καθάρισμα σε στοιχεία emails, Αριθμών Δελτίων Ταυτότητας και Διαβατηρίων και εφαρμογή πρόσθετων business κανόνων και εμπλουτισμός υφιστάμενων διαδικασιών ομαδοποίησης εγγραφών CAs σε Golden Record Groups.

Για τον λόγο αυτό η ΔΕΗ επιθυμεί να προσλάβει εξειδικευμένο Σύμβουλο για την υλοποίηση των τεχνικών αυτών εργασιών στην υφιστάμενη υποδομή της ΔΕΗ, δηλαδή MS Azure cloud με χρήση εργαλείων Information Server και Mnemosyne. Αναλυτικότερα, ο Σύμβουλος θα αναλάβει τον σχεδιασμό, την υλοποίηση και την συντήρηση των ροών εκτέλεσης των παραπάνω τεχνικών εργασιών καθώς και την υποστήριξη της αρμόδιας ομάδας της ΔΕΗ (Διεύθυνσης Στρατηγικής Δεδομένων & Ανάλυσης) στις περιόδους μεταφόρτωσης δεδομένων στην νέα πλατφόρμα πελατοκεντρικής εξυπηρέτησης CRM.

## Στοιχεία Παρεχόμενης Υπηρεσίας

### Γενικά

Στα πλαίσια του έργου, ο Ανάδοχος Σύμβουλος θα παραλάβει ένα dataset με την ακόλουθη master πληροφορία πελατειακής βάσης:

- ✓ Αριθμός Συμβολαίου (CA)
- ✓ Λεκτικά/Ονοματεπώνυμα Αντι-συμβαλλόμενων ή Company Names
- ✓ Πλήρη στοιχεία Διευθύνσεων B2B, Αποστολής Λογαριασμών και Μετρητών/Παροχών
- ✓ Τηλέφωνα επικοινωνίας
- ✓ Email επικοινωνίας
- ✓ Ημερομηνίες ένταξης & απένταξης συμβολαίων
- ✓ ΑΦΜ και ΔΟΥ
- ✓ Αριθμοί Δελτίων Ταυτότητας
- ✓ Αρ. Διαβατηρίου

καθώς και την υφιστάμενη ροή ομαδοποίησης σε επίπεδο πελάτη (φυσικού και νομικού προσώπου) που συμπληρώνει τα παραπάνω με το εξής πεδίο:

### Golden Record ID

και τα αντίστοιχα Golden Record Group fields (δηλαδή Λεκτικά Ονοματεπώνυμα και Company Name, Διευθύνσεις αποστολής λογαριασμού και έδρας, τηλέφωνα επικοινωνίας και emails).

Στην υφιστάμενη ροή προβλέπεται η δυνατότητα τακτικής εβδομαδιαίας ανανέωσης στο κατάλληλα διαμορφωμένο production περιβάλλον του MS Azure cloud της ΔΕΗ.

Η αναμενόμενη υπηρεσία από τον εξωτερικό συνεργάτη θα περιλαμβάνει την τροποποίηση της υφιστάμενης ροής με τρόπο που να συμπεριλαμβάνει και τις ακόλουθες πρόσθετες προδιαγραφές, όπως αναφέρονται λεπτομερώς ακολούθως.

Ειδικότερα, η διαδικασία που θα ακολουθείται για την εκτέλεση στο παραγωγικό περιβάλλον κάθε αναγκαίας τροποποίησης είναι η εξής: Κάθε τροποποίηση θα υλοποιείται, και τεστάρεται στο κατάλληλα διαμορφωμένο staging περιβάλλον του MS Azure cloud της ΔΕΗ και μόνο μετά από ενδελεχείς UAT ελέγχους από την αρμόδια ομάδα της ΔΕΗ και το σχετικό green light θα περνούν οι εγκεκριμένες υλοποιημένες ροές στο κατάλληλα διαμορφωμένο production περιβάλλον του MS Azure cloud της ΔΕΗ.

#### Λεπτομερείς Προδιαγραφές

1. Προσθήκη Φύλου Αντι-Συμβαλλόμενου (Gender): Η προσθήκη του νέου πεδίου θα πρέπει να λάβει υπόψη του: (i). τα λεξιλόγια γυναικείων και ανδρικών ονομάτων ελληνικής γλώσσας (με ελληνικούς και λατινικούς χαρακτήρες) καθώς και λεξιλόγια ξένων γυναικείων και ανδρικών ονομάτων (με λατινικούς χαρακτήρες), (ii). την κατάληξη του επωνύμου και (iii). τον βαθμό βεβαιότητας του αντίστοιχου υπολογισμού με στόχο η αρμόδια ομάδα της ΔΕΗ να μπορεί να αξιολογήσει την αξιοπιστία του αποτελέσματος και συνακόλουθα να είναι σε θέση να λάβει τεκμηριωμένη απόφαση για την τελική χρήση και μεταφόρτωση της πληροφορίας στην πλατφόρμα πελατοκεντρικής εξυπηρέτησης CRM.

2. Αναγνώριση και διόρθωση email: Στα email addresses που έχουν ήδη χαρακτηριστεί ως *invalid* θα εφαρμοσθεί περαιτέρω ανάλυση και επεξεργασία ώστε να: (i). προσδιορισθεί email μέσα σε κείμενο (π.χ. <dimitris arampatzis> dimisaram@gmail.com), (ii). συμπληρωθεί email (π.χ. apadopoulosstamatis@gmail) και (iii). διορθωθούν τυχόν τυπογραφικά σφάλματα (π.χ. thanasia.20@hotmail.con)

3. Αναγνώριση και διόρθωση AT (στρατιωτικής, αστυνομικής, πολιτικής) & Διαβατήριο (Ελληνικό ή Ξένο): Τα πεδία αριθμού ταυτότητας, αριθμού διαβατηρίου και χώρα προέλευσης («INSTITUTE») θα αξιοποιηθούν με κατάλληλο τρόπο ώστε να πραγματοποιηθούν όλες οι αναγκαίες εργασίες για να: (i) εντοπισθεί το κατάλληλο format, (ii). αναγνωρισθεί και (iii). αποθηκευτεί το ακριβές identifier. Ενδεικτικά παραδείγματα τέτοιου είδους εργασιών αποτελούν:

a. τυχόν αναγκαία μετατροπή χαρακτήρων από ελληνικούς σε λατινικούς και αντίστροφα,

b. εφαρμογή regular expressions με λεκτικά στρατιωτικών, αστυνομικών, πολιτικών ταυτοτήτων (όπως για παράδειγμα ΓΕΣ, ΓΕΝ, ΓΕΑ, ΓΕΕΘΑ, ΕΛΑΣ, κτλ.)

c. εφαρμογή κατάλληλων regular expressions για εντοπισμό διαβατηρίων, ανάλογα με την χώρα προέλευσης

Το τελικό αποτέλεσμα αυτής της διαδικασίας θα είναι η τροποποίηση του υφιστάμενου πεδίου AT\_Info με τις ακόλουθες διακριτές τιμές:

d. AT\_Format

e. Invalid\_Format

f. Passport\_Format

g. Army\_Police\_Format

h. Null

καθώς και ο σχετικός εμπλουτισμός του πεδίου CL\_Passport με τα επιπλέον διαβατήρια που θα προκύψουν από τις παραπάνω πρόσθετες τεχνικές εργασίες.

4. Αναγνώριση είδους πελάτη (customer type): Στην υφιστάμενη υλοποίηση το είδος πελάτη έχει αναγνωρισθεί με βάση το αρχικό ψηφίο του VAT και συγκεκριμένα:

a. If starting\_digit(VAT\_ID) = 1 – 6 then Customer\_Type προκύπτει από τα λεκτικά Ονοματεπωνύμου (λαμβάνοντας ως πιθανές τιμές person, organization/company, koinoxrista, municipality, other)

b. If starting\_digit(VAT\_ID) = 7 – 9 then Customer\_Type=organization

c. If VAT\_ID = null or invalid then Customer\_Type = null

Στις περιπτώσεις που το VAT είναι αρχικά κενό αλλά στην συνέχεια έχει συμπληρωθεί μέσω του Golden Record Group οι παραπάνω κανόνες χρησιμοποιούνται για το ΑΦΜ που έχει βρεθεί μέσω Golden Record Group.

Στο σημείο αυτό αναμένονται από τον εξωτερικό συνεργάτη πρόσθετες προτάσεις κανόνων/best practices για περαιτέρω βελτίωση της ορθής αναγνώρισης τύπου πελάτη στο πελατολόγιο της ΔΕΗ και με βάση τις προηγούμενες σχετικές εμπειρίες.

5. Αναγνώριση ορθότητας ευρέως χρησιμοποιούμενων ΑΦΜ με βάση τα λεκτικά (πχ. ΔΕΗ/ΕΥΔΑΠ, κτλ): Στην πελατειακή βάση της ΔΕΗ ένα πολύ μεγάλο πλήθος CAs έχουν αντιστοιχισθεί εσφαλμένα στο ΑΦΜ της ΔΕΗ, ΕΥΔΑΠ, κ.τ.λ. πράγμα που θα πρέπει να αντικατασταθεί με null VAT και στην συνέχεια να ακολουθηθεί η ίδια διαδικασία χειρισμού των υπόλοιπων περιπτώσεων κενών ΑΦΜ. Πιο συγκεκριμένα θα πρέπει να εντοπισθεί το υποσύνολο των περιπτώσεων που στα λεκτικά εμφανίζονται σχετικοί όροι (πχ. ΔΕΗ, ΕΥΔΑΠ, κτλ) και μόνο σε αυτές να θεωρηθεί το ΑΦΜ έγκυρο. Στις υπόλοιπες θα πρέπει να γίνει η προ-αναφερόμενη αντικατάσταση με null ΑΦΜ.

6. Αναγνώριση και απομόνωση εγγραφών CAs με συγκεκριμένα λεκτικά: Λεκτικά όπως «Χωρίς Προμηθευτή», «Κομμένο» κ.τ.λ. στην πελατειακή βάση της ΔΕΗ δηλώνουν ανάγκη για εξαίρεση. Για τον λόγο αυτό η αρμόδια ομάδα της ΔΕΗ θα προσδώσει την σχετική λίστα εξαιρέσεων (υπό την μορφή keywords) στον εξωτερικό συνεργάτη ώστε να αναγνωρίσει, απομονώσει και εξαιρέσει από τις συνακόλουθες τεχνικές διαδικασίες τις αντίστοιχες εγγραφές.

7. Αντικατάσταση κενών Billing Addresses με POD Addresses: Στην πελατειακή βάση της ΔΕΗ υπάρχουν ~599.9K CAs με κενά Billing Addresses, εκ των οποίων τα ~599.8K έχουν συμπληρωμένα τα αντίστοιχα πεδία POD Addresses. Για τις τελευταίες περιπτώσεις θα πρέπει τα κενά Billing Addresses να αντικατασταθούν με τα PoD Addresses και να εφαρμοστούν όλες οι υπόλοιπες διαδικασίες Golden Record ομαδοποίησης στα εμπλουτισμένα δεδομένα διευθύνσεων.

8. Κανονικοποίηση ονομάτων B2B\_Company\_Name: Θα πρέπει να εφαρμοστεί κανονικοποίηση (κοινή γραφή) των B2B\_Company\_Name για τις κοινότητες/δήμους με γνωστά ονόματα.

9. Ξεχωριστή υλοποίηση για Golden Record Groups τύπου ΠΟΛΛΑΠΛΩΝ: Οι «ΠΟΛΛΑΠΛΟΙ» πελάτες στην ορολογία της ΔΕΗ αφορούν τιμολόγηση, μαζικές

επικοινωνίες, sms/viber καμπάνιες, κτλ. και μπορεί να είναι «ενταγμένοι» στην υφιστάμενη πελατειακή βάση της ΔΕΗ ή όχι και γι' αυτό χρίζουν ειδικής μεταχείρισης. Ειδικότερα οι λεπτομέρειες της σχετικής αναγκαίας υλοποίησης ακολουθούν κατά περίπτωση:

#### A. Ιδιώτες

Ο «πελάτης» ορίζεται ως:

- i. CAs που έχουν ίδιο Ar\_Πολλαπλού
- ii. CAs με κενό Ar\_Πολλαπλού και ΑΦΜ ίδιο με το «Βασικό» ΑΦΜ του υπό εξέταση Ar\_Πολλαπλού, όπως εντοπίζεται στο πεδίο *VKTyp* του SAP ISU πίνακα *FKKVK*

#### B. Δημόσιο

Θα δημιουργηθεί μια ξεχωριστή κατηγορία Customer\_Type\_DImosio=Y, που όμως θα λαμβάνει τιμές "Y" μόνο για το σχετικό υποσύνολο των εγγραφών του customer\_type και θα προέρχεται από το 2<sup>ο</sup> ψηφίο του Εμπορικού Εταίρου = 4 και πιο συγκεκριμένα το πεδίο VBUND στον πίνακα FKKVKP.

Ο βασικός ορισμός «πελάτη» παραμένει ο ίδιος με παραπάνω, δηλαδή:

- i. CAs που έχουν ίδιο Ar\_Πολλαπλού
- ii. CAs με κενό Ar\_Πολλαπλού και ΑΦΜ ίδιο με το «Βασικό» ΑΦΜ του υπό εξέταση Ar\_Πολλαπλού, όπως εντοπίζεται στο πεδίο *VKTyp* του SAP ISU πίνακα *FKKVK*

#### Γ. Δήμοι – Περίπτωση I

Η κατηγορία «ΔΗΜΟΙ» εντοπίζεται από το Excel «Δήμοι Καλλικράτη» στην βάση του Ar\_Πολλαπλού (Στο προαναφερόμενο Excel αναφέρεται ως "ΛΟΓ. ΣΥΜΒ. SAP". Πιο συγκεκριμένα ο «πελάτης» είναι η κάθε κοινότητα του κάθε Δήμου (δηλαδή η ξεχωριστή εγγραφή του συνημμένου Excel με τους Καλλικρατικούς Δήμους:

- i. Όλες οι κοινότητες / διαφορετικοί Ar\_Πολλαπλού όπως εμφανίζονται στο συνημμένο Excel και τα διακριτά CAs που ανήκουν στον κάθε ένα από αυτούς τους πολλαπλούς
- ii. CAs με κενό Ar\_Πολλαπλού και ΑΦΜ ίδιο με του κάθε Δήμου στον οποίο αντιστοιχεί ο κάθε Ar\_Πολλαπλού

Πιο συγκεκριμένα η διασύνδεση πρέπει να γίνει πάνω στον Ar.Πολλαπλού (ονομασία «ΛΟΓ.ΣΥΜΒ SAP"- στήλη B Excel Καλλικράτη) και θα πρέπει να συμπεριλαμβάνονται στον τελικό πίνακα οι στήλες C – Κοινότητα, D – κωδικός Καλλικράτη και K – Δήμος Καλλικράτη από το Excel «Δήμοι Καλλικράτη». Στο σημείο αυτό αποσαφηνίζεται ότι η πληροφορία της στήλης K – Δήμος Καλλικράτη ουσιαστικά θα υλοποιείται μέσω της κατηγορίας «Πολλαπλοί Πολλαπλοί» που ακολουθεί.

### Γ. Δήμοι – Περίπτωση II

Όσα δεν έχουν Αρ\_Πολλαπλού ούτε ΑΦΜ που ανήκει σε Πολλαπλό θα αναγνωριστούν με fuzzy matching όλα τα λεκτικά που περιλαμβάνουν κοινότητες κι έτσι να ενταχθούν σε διαφορετικά Golden Record Groups και customer\_type=ΔΗΜΟΙ. Για τις περιπτώσεις όπου στο όνομα υπάρχει επιπλέον πληροφορία εκτός από το όνομα Δήμου/Κοινότητα, (πχ. ΔΗΜΟΣ ΠΕΡΙΣΤΕΡΙΟΥ ΑΝΟΙΚΤ.ΚΟΛΥΜΒΗΤΗΡ., ΔΗΜΟΣ ΧΑΛΚΙΔΑΣ PARKING ΔΗΜ. ΑΓΟΡΑΣ, ΒΡΕΦΟΝΗΠΙΑΚΟΣ ΣΤΑΘΜΟΣ ΔΗΜΟΥ ΑΛΙΜΟΥΚΟΙΝΟΤΙΚΟ ΑΝΤΛΙΟΣΤ ΝΑΟΥΣΑΣ ΠΑΡΟΥ, ΔΗΜΟΤΙΚΟ ΑΝΤΛΙΟΣΤΑΣΙΟ...) το company\_type θα χαρακτηριστεί ως ΔΗΜΟΙ/municipality.

Σημειώνεται ότι στις περιπτώσεις αυτές δεν θα υπάρχει αντίστοιχος Κωδικός Καλλικρατικού Δήμου, καθώς η περίπτωση αυτή δεν εντάσσεται ούτε στους «Ενταγμένους» ούτε στους «Ανένταχτους».

Τέλος οι περιπτώσεις που δεν αναφέρουν τον όνομα Δήμου αλλά 'παραπέμπουν' σε ΔΗΜΟ (πχ. ΔΗΜΟΤΙΚΟ ΑΝΤΛΙΟΣΤΑΣΙΟ, ΚΟΙΝ.ΚΑΜΠΙΓΚ, ΚΟΙΝΟΤΙΚΗ ΑΠΟΧΕΤΕΥΣΗ, ΦΟΠ ΠΛΑΤΕΙΑΣ ΚΕΡΑΣΙΑ, ΔΗΜΟΤΙΚΟ ΣΧΟΛ. 7, 1 ΔΗΜΟΤΙΚΟ ΣΧΟΛΕΙΟ ... ) δεν θα συμπεριλαμβάνονται στην κατηγορία customer\_type=ΔΗΜΟΙ/municipality.

### Δ. Πολλαπλοί «Πολλαπλοί»

Για τις περιπτώσεις που περισσότεροι από 1 αριθμοί πολλαπλών αντιστοιχούν στον ίδιο «πελάτη» θα δημιουργηθεί ξεχωριστό πεδίο που να ομαδοποιεί τους πολλαπλούς Αριθμούς\_Πολλαπλών με το όνομα Multiple\_Pollaplos. Το σχετικό groupάρισμα θα γίνει με όμοια διαδικασία fuzzy matching στα λεκτικά των επωνυμιών, όπως γίνεται στις υπόλοιπες περιπτώσεις Golden Record group.

10. Αναγνώριση περιπτώσεων Πολλαπλών με «λάθος καταχωρημένο ΑΦΜ»: Για τις περιπτώσεις «Μη ενταγμένων» Πολλαπλών με χρήση Βασικού ΑΦΜ Πολλαπλού θα πρέπει να γίνει εξαντλητικός (exhaustive) έλεγχος ορθής αντιστοίχισης Βασικού ΑΦΜ και λεκτικών ονοματεπωνύμων αντίστοιχων εγγραφών Contract Account. Για τις περιπτώσεις που εντοπίζονται αποκλίσεις τα βασικά ΑΦΜ θα πρέπει να διορθωθούν/ακυρωθούν κατάλληλα σε συμφωνία με την αρμόδια ομάδα της ΔΕΗ.

11. Δημιουργία πεδίου «Μέθοδος Ομαδοποίησης» ειδικά για την περίπτωση Πολλαπλών ώστε να διαφοροποιείται η μέθοδος ομαδοποίησης των CAs σε κάθε Golden Record Group: Το πεδίο «Μέθοδος Ομαδοποίησης» μπορεί να παίρνει τις τιμές: «Ενταγμένος», «Ανένταχτος» και «Αλγόριθμος». Η τιμή «Ενταγμένος» προέρχεται από ΑΦΜ ή Αρ\_Πολλαπλού, η τιμή «Ανένταχτος» από τους ανένταχτους πολλαπλούς όπως περιγράφονται παραπάνω (δηλαδή «βασικό» ΑΦΜ Πολλαπλού) και η τιμή «Αλγόριθμος» από τα matching criteria & fuzzy matching τεχνικές, όπως περιεγράφηκαν παραπάνω.

12. Ένταξη στην διαδικασία matching για Golden Record Group Analysis πεδίων σχετικών με Ebill: Η αναγνώριση Golden Record Groups θα πρέπει να συνοψολογίζει και τα πεδία Ebill\_User\_Name και Ebill\_Email (που δεν χρησιμοποιούνται στην υφιστάμενη διαδικασία).

13. Ένταξη πεδίων σχετικών με Ταυτότητες, Διαβατήρια & E-bill στα Golden Record Group – level πεδία: Ανάμεσα στα υπόλοιπα πεδία επιπέδου Golden Record Group που ήδη περιλαμβάνονται στο υφιστάμενο τελικό output dataset που θα

παραλάβει ο εξωτερικός ανάδοχος (δηλ. Λεκτικά Ονοματεπώνυμα και Company Name, Διευθύνσεις αποστολής λογαριασμού και έδρας, τηλέφωνα επικοινωνίας και emails) θα πρέπει να προστεθούν:

a. πεδία ταυτοτήτων AT, διαβατηρίων Passport όπως περιεγράφηκαν παραπάνω και συγκεκριμένα και τα πεδία ισχύος των εν λόγω εγγράφων (δηλ. ID\_VALID\_DATE\_FROM, ID\_AT\_INSTITUTE, Passport\_VALID\_DATE\_TO)

b. πεδία που σχετίζονται με το Ebill account (δηλ. GR\_Ebill\_Activation\_Date, GR\_Ebill\_User\_Name, GR\_Password, GR\_Ebill\_Delete\_Date, GR\_Ebill\_Transformed\_Email).

14. Δημιουργία κοινής μεθοδολογίας συμπλήρωσης πεδίων επιπέδου Golden Record Group από τα υφιστάμενα πεδία επιπέδου CA: Τα διαφορετικά πεδία επιπέδου Golden Record Group θα πρέπει να προέρχονται από την πιο συχνή τιμή κάθε αντίστοιχου πεδίου από τα διαφορετικά CAs που ανήκουν στο ίδιο Golden Record Group. Μοναδική εξαίρεση αποτελούν τα πεδία που σχετίζονται με Ταυτότητες και Διαβατήρια και το Ebill για τα οποία θα χρησιμοποιούνται η πιο πρόσφατη ημερομηνία έκδοσης (αντίστοιχα πεδία: ID\_VALID\_DATE\_FROM και Ebill\_Activation\_Date).

15. Δημιουργία GR\_Customer\_Type που θα αφορά το είδος πελάτη του Golden Record Group και όχι του CA (σε επίπεδο Golden Record Group)

16. Υλοποίηση Extra post-processing requirements: Θα χρειαστεί να υλοποιηθούν κάποιοι πρόσθετοι κανόνες ομαδοποίησης CAs ή/και τροποποίησης της αλγοριθμικής/αυτόματης διαδικασίας ένταξης σε Golden Record groups με βάση business κανόνες που θα προδιαγραφούν από την αρμόδια ομάδα της ΔΕΗ ή/και τον οπτικό έλεγχο των αποτελεσμάτων που θα προκύψουν κατά την διάρκεια του έργου

17. Υποστήριξη για τις ανάγκες μεταφόρτωσης των δεδομένων στην νέα CRM πλατφόρμα: Στα πλαίσια της υποστήριξης της διαδικασίας προετοιμασίας και μεταφόρτωσης των δεδομένων στην νέα CRM πλατφόρμα του οργανισμού (Salesforce), ο εξωτερικός ανάδοχος θα χρειαστεί να υποστηρίξει την αρμόδια ομάδα της ΔΕΗ για σταδιακές φορτώσεις και επανα-φορτώσεις δεδομένων, παράδοση reconciliation reports και φυσικά υποστήριξη και συμμετοχή κατά το χρονικό διάστημα μετάβασης στο SF.

18. Παροχή score ομοιότητας εγγραφών CAs που έχουν ομαδοποιηθεί στο ίδιο Golden Record Group σε κοινή κλίμακα αξιολόγησης: Στόχος του εν λόγω score, που εφεξής καλούμε Composite Similarity Score, είναι η αποτίμηση του ποσοστού βεβαιότητας της κατάταξης ή μη ένταξης κάθε εγγραφής σε ένα Golden Record group προκειμένου για την τεκμηριωμένη αξιολόγηση από την αρμόδια ομάδα της ΔΕΗ. Στο απαιτούμενο score ομοιότητας θα πρέπει να συνυπολογιστούν κατ'ελάχιστον:

a. το είδος της κατάταξης της κάθε εγγραφής (δηλ. Main Record, Duplicate, Clerical, Residual),

b. το πλήθος των απαιτούμενων passes για την τελική κατάταξη ή/και μη ένταξη,



c. ο τύπος του (-ων) πεδίου(-ων) που χρησιμοποιείται για blocking ή/και matching criteria (πχ λεκτικά Ονοματεπώνυμα, VAT, ΑΔΤ, Αρ. Διαβατηρίου, Διευθύνσεις, τηλέφωνα επικοινωνίας, email, customer type, κτλ.),

d. το αν το κάθε ένα από αυτά τα πεδία είναι κενό ή περιέχει αξιοποιήσιμη πληροφορία

Για το εν λόγω score ομοιότητας η ομάδα της ΔΕΗ αναμένει από τον εξωτερικό συνεργάτη να προτείνει την ακριβή μεθοδολογία που ανταποκρίνεται με τον καλύτερο δυνατό τρόπο στα δεδομένα της ΔΕΗ. Ενδεικτικά παραδείγματα αποτελούν η χρήση των σταθμίσεων ομοιότητας που παρέχονται από την πλατφόρμα, distance functions όπως η Levensthein distance ή/και η ομοιότητα συνημιτόνου (cosine similarity), παραλλαγές αυτών ή/και κατάλληλοι συνδυασμοί μεταξύ τους.

19. Βελτίωση Ομαδοποίησης CAs και με βάση τεχνικές διαδικασίες fuzzy matching λεκτικών πεδίων CAs: Θα πρέπει να υλοποιηθούν στατιστικές και λεξικογραφικές μέθοδοι ομοιότητας σε κάθε πιθανό ζεύγος «όμοιων» εγγραφών. Για τον σκοπό αυτό προϋποτίθεται η εκτέλεση όλων των τυχόν αναγκαίων κανονικοποιήσεων σε όρους που συναντιούνται συχνά στα δεδομένα της ΔΕΗ, όπως για παράδειγμα λεκτικά που αφορούν Κοινόχρηστα, Δήμους, Δημοτικά Σχολεία, Αντλιοστάσια, Στρατόπεδα, κτλ). Στην υλοποίηση της ομαδοποίησης με χρήση τεχνικών fuzzy matching λεκτικών θα πρέπει να συμπεριληφθεί και το αντίστοιχο confidence level (επίπεδο σιγουριάς) ομοιότητας κάθε πιθανού ζεύγους, που με την σειρά του θα συνεκτιμά ανάμεσα σε άλλους πιθανούς παράγοντες και το πλήθος αναγνωριστικών πεδίων που χρησιμοποιήθηκαν για τον σκοπό αυτό καθώς και οι κατάλληλες σταθμίσεις. Ο εξωτερικός συνεργάτης αναμένεται να περιγράψει και υλοποιήσει την ακριβή προτεινόμενη για τον σκοπό αυτό μεθοδολογία. Τα αποτελέσματα θα συγκριθούν με τα υφιστάμενα, ώστε να διασφαλιστεί από την αρμόδια ομάδα της ΔΕΗ η σχετική βελτίωση που θα επέλθει από το αντίστοιχο παραδοτέο.

20. Δημιουργία αυτόματου μηχανισμού delta επικαιροποίησης: Η διαδικασία αυτή θα πρέπει να περιλαμβάνει τις αλλαγές που έχουν πραγματοποιηθεί στα δεδομένα ανάμεσα στα δυο διαδοχικά runs, δηλαδή: (i). πιθανές τροποποιήσεις σε CAs που υπήρχαν στο προηγούμενο run και είχαν κατηγοριοποιηθεί ως Master, Duplicates, Clerical ή Residuals και είχαν αντίστοιχα συμπεριληφθεί ή όχι σε κάποιο Golden Record Group και (ii). νέα συμβόλαια CAs. Στο σημείο αυτό αποσαφηνίζεται ότι οι πιθανές τροποποιήσεις σε υφιστάμενα συμβόλαια όπως περιγράφονται στο (i) παραπάνω ενδέχεται να προκαλέσουν διαφορετικό χαρακτηρισμό των εν λόγω CAs κι επομένως απένταξη ή διαφορετική ένταξη σε Golden Record Groups. Τα νέα συμβόλαια όπως αναφέρονται στο (ii) παραπάνω θα πρέπει να συγκριθούν με την προτεινόμενη μεθοδολογία "Composite Similarity Score" με όλα τα Golden Record Groups προκειμένου να εντοπισθεί η ορθότερη ένταξη τους. Σε κάθε κύκλο τακτικής (εβδομαδιαίας) ανανέωσης ο πλήρης μηχανισμός καθαρισμού δεδομένων και Golden Record Analysis θα πρέπει να εφαρμόζεται μόνο σε όσα από τα records/CAs πληρούν τις παραπάνω προδιαγραφές, που θα πρέπει να αναγνωρίζονται με έναν γρήγορο και εύκολο τρόπο μέσω κατάλληλου flaggarίσματος.

21. Διατήρηση σταθερού Golden Record ID μεταξύ διαφορετικών runs του μηχανισμού τακτικής (εβδομαδιαίας) ανανέωσης: Ανάμεσα στα διαδοχικά runs τα Golden Record Group identifiers πρέπει να παραμένουν σταθερά προκειμένου να



μπορεί να δουλέψει ο αυτόματος μηχανισμός delta επικαιροποίησης, όπως περιγράφεται παραπάνω. Αυτό σημαίνει ότι το identification των Golden Record Groups απαιτείται να είναι ανεξάρτητο από τα ακριβή CA-μέλη του, καθώς πιθανές τροποποιήσεις των στοιχείων κάποιου/-ων από αυτά ενδέχεται να τα αλλάξουν ανάμεσα σε διαδοχικά runs και μόνο με την διατήρηση της σταθερότητας για τα εναπομείναντα μέλη θα μπορεί να επιτευχθεί η ορθότητα των υπολογισμών σύμφωνα με την προτεινόμενη μέθοδο "Composite Similarity Score" για την ορθή ένταξη των νέων CAs σε κάποιο από τα ήδη υπάρχοντα Golden Record groups (που έχουν δημιουργηθεί είτε στο προηγούμενο run, ακόμα κι αν τα ακριβή μέλη τους μπορεί να έχουν μεταβληθεί στο νέο run, είτε να έχουν δημιουργηθεί εξ'ολοκλήρου στο νέο run).

#### Παραδοτέα – Διάρκεια Έργου

Ο Αντισυμβαλλόμενος υποχρεούται να παραδώσει τα ακόλουθα:

- 1) Λεπτομερές χρονο-διάγραμμα του έργου
  - 2) Περιγραφή του ακριβούς τρόπου υλοποίησης δημιουργίας Golden Record Group, δηλαδή αν και σε ποιον βαθμό θα αξιοποιείται η υφιστάμενη υλοποίηση ή αν θα πραγματοποιηθεί επανα-υλοποίηση με αντίστοιχη τεκμηρίωση της κατ' ελάχιστον επίτευξης του υφιστάμενου βαθμού αξιοπιστίας
  - 3) Σε περίπτωση και για όσα υποσύνολα της υφιστάμενης λειτουργικότητας Golden Record επανα-υλοποιηθούν στα πλαίσια αυτού του έργου: output dataset που περιλαμβάνει τα αντίστοιχα πεδία που επανα-υλοποιήθηκαν ώστε να μπορούν να ελεγχθούν ως προς την προβλεπτική τους ικανότητα κι έτσι να επιβεβαιωθεί από την αρμόδια ομάδα της ΔΕΗ η αναγκαία κατ' ελάχιστον επίτευξη της υφιστάμενης προβλεπτικής ικανότητας
  - 4) Τελικό consolidated output dataset με τις πρόσθετες προδιαγραφές που περιγράφονται παραπάνω στο αποτέλεσμα των πιο πρόσφατων επικαιροποιημένων στοιχείων βάσης δεδομένων της ΔΕΗ
  - 5) Ροές που υλοποιούν τις παραπάνω προδιαγραφές φορτωμένες και στο staging και production περιβάλλον στο MS Azure της ΔΕΗ σύμφωνα με τις παραπάνω διαδικασίες
  - 6) Κάθε τροποποίηση που υλοποιείται πρέπει να περνά στο staging περιβάλλον προς UAT έλεγχο από την αρμόδια ομάδα της ΔΕΗ και μόνο μετά από το green light να περνά στο production περιβάλλον
  - 7) Τεκμηρίωση και περιγραφή της μεθοδολογίας και του κώδικα/ρωών που θα παραδοθούν, καθώς και training της αρμόδιας ομάδας της ΔΕΗ
- Η ΔΕΗ επιθυμεί το πλάνο ολοκλήρωσης του έργου να μην υπερβαίνει συνολικά τις πέντε (5) εργάσιμες εβδομάδες. Ειδικότερα η ΔΕΗ επιθυμεί από την ημερομηνία πρόσδοσης προσβάσεων και παράδοσης της βάσης δεδομένων ο ανάδοχος σύμβουλος να έχει διεκπεραιώσει όλα τα παραδοτέα το αργότερο εντός πέντε (5) εργάσιμων εβδομάδων. Επίσης, κατά την διάρκεια της πρώτης (1ης) εργάσιμης εβδομάδας ο ανάδοχος υποχρεούνται να παραδώσει το λεπτομερές χρονοδιάγραμμα εργασιών του έργου και την λεπτομερή αρχιτεκτονική της προτεινόμενης υλοποίησης, που αποτελούν τα παραδοτέα νούμερο 1 & 2 του έργου, αντιστοίχως. Επιπλέον διευκρινίζεται ότι στο παραδοτέο 2 θα αποσαφηνίζεται ο ακριβής τρόπος υλοποίησης δημιουργίας Golden Record Group, δηλαδή αν και σε ποιον βαθμό θα αξιοποιείται η υφιστάμενη υλοποίηση ή αν θα πραγματοποιηθεί επανα-υλοποίηση με αντίστοιχη τεκμηρίωση της κατ' ελάχιστον επίτευξης του υφιστάμενου βαθμού αξιοπιστίας. Σε περίπτωση που προκύψει αναγκαία επανα-υλοποίηση μέρους ή του συνόλου της υφιστάμενης λειτουργικότητας απόδοσης Golden Record Group μέχρι το τέλος της 2ης εβδομάδας του έργου θα πρέπει να έχει παραδοθεί η σχετική απαιτούμενη επανα-υλοποίηση με τεκμηρίωση της κατ' ελάχιστον επίτευξης του υφιστάμενου βαθμού προβλεπτικής ικανότητας (Παραδοτέο 3). Για την τεκμηρίωση αυτή η αρμόδια ομάδα της ΔΕΗ θα πραγματοποιήσει τους σχετικούς ελέγχους επί του Παραδοτέου 3.



Επιπλέον, αποτελεί υποχρέωση του συμβούλου να παρέχει υποστήριξη και μετά την ολοκλήρωση του έργου και για χρονικό διάστημα τουλάχιστον ενός (1) μήνα μετά στην αρμόδια ομάδα της ΔΕΗ αναφορικά με επεξηγήσεις και διευκρινίσεις που πιθανόν να απαιτηθούν για τα παραδοτέα, καθώς και στις συγκεκριμένες ανάγκες σταδιακών φορτώσεων δεδομένων στην νέα πλατφόρμα SF και στο τελικό χρονικό διάστημα μεταφόρτωσης δεδομένων πριν από το εμπορικό λανσάρισμα της νέας πλατφόρμας SF.